



LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

THE SIZE DISTRIBUTION OF OIL AND GAS FIELDS

by

Gordon M. Kaufman

June 1964

62-64

THE SIZE DISTRIBUTION OF OIL AND GAS FIELDS

1. Introduction

Statistical decision theory is the mathematical analysis of decision problems in which uncertainty is a key element. Use of it can give the oil and gas operator ordered insight into the way exploration decisions should be made in order to achieve preassigned goals; e.g. maximize the net expected present value of an exploratory drilling program. It does this by forcing the decision-maker to make explicit assumptions and judgements that are implicit in every such decision problem. In other words, statistical decision theory is a vehicle for rendering precise the key variables and relations among them that constitute the core of exploration problems. It is important to recognize, however, that people--not mathematical models--make decisions, so statistical decision theory is only an aid to decision-making, not a decision-making device.

Rather than emphasize the specifics of how statistical decision theory comes into play in analysis of exploration problems, in this paper we will concentrate on illustrating how statistical methodology can be used to build a probabilistic cornerstone of mathematical models of some important exploration decision problems.

As pointed out in [1], in any decision concerned with the strategy and tactics of oil and gas exploration, a key variable is the size of hydrocarbon deposits in barrels of oil or in MCF of gas. The size of pool or field[†] discovered in a particular wildcat venture determines the degree

[†]For convenience we shall call a hydrocarbon deposit a "pool" or a "field", even though the terms differ in usage in that a new "pool" can be found within an already discovered "field". By convention, "size" will always refer to size in barrels of oil or in MCF of gas. The word "area" will denote areal extent.

to which the venture is an economic success. Since the pool or field size that will be discovered is almost always unknown before a prospect is drilled, an important question is:

What functional form of distribution function should be used to characterize the probability distribution of field sizes in a petroleum province?

By "functional form" we mean a mathematical formula which defines a family of distribution functions.

Clearly the functional form used to characterize the size distribution of oil and gas fields is a vital part of any model which you as decision makers might use to analyze exploration decisions. Ideally, we would like this form to be flexible enough to fit a wide variety of empirical histograms of oil and gas fields in differing areas with differing definitions of reserves by varying only the value of the parameters of the form, not the form itself. We also would like it to be analytically tractable, so that it may be easily used in the course of a formal analysis of exploration decision problems; e.g. by use of statistical decision models. The Lognormal functional form has these properties.[†]

In addition to possessing these properties the Lognormal distribution has other desirable attributes;

1. it may be shown to be in concordance with some concepts of how mineral deposits are formed;

[†]A random variable is said to be "Lognormally distributed", if the logarithm of the random variable is Normal or Gaussian; cf. Appendix A.

2. stochastic models of the discovery process built on reasonable assumptions about the process lead to the Lognormal functional form;
3. the class of Lognormal functional forms is analytically tractable and flexible enough to capture most reasonable oilmen's subjective betting odds about random variables such as reported field size.

A detailed discussion of these points, a development of several methods for blending subjective probability beliefs of experts about field sizes that are not known with certainty with objective evidence, and application of these methods to some typical exploration decision problems--notably drilling decision problems--are given in [1] and [2]. Both references show how concepts from statistical decision theory can be used to analyze such problems.

My purpose here is three-fold: first, to show that the Lognormal distribution provides a reasonable fit to empirical histograms of reported oil fields sizes, where "size" is measured in barrels of ultimate primary recoverable reserves; second, to illustrate how we can use the Lognormal distribution to describe systematically one important dimension of the discovery process--the build-up over time of such histograms; and third, to demonstrate with a hypothetical example how we can use properties of the Lognormal distribution as an aid in assessing how the probability distribution of field sizes remaining to be discovered in a basin varies with time. Knowing properties of this last mentioned probability distribution is extremely useful in assessing the effect of time of entry on the expected profitability of a major exploration program in a given basin.

2. Empirical Histograms of Reported Oil Field Sizes

2.1 Data Sources

Oil and Gas Journal statistics[†] on reported field sizes are plotted in Exhibits 1 through 9 to demonstrate how the Lognormal functional form can be fitted to empirical histograms of reported oil field sizes and to show how we may characterize the change over time of these histograms using only a pair of numbers--the parameters of the Lognormal distribution--for each point in time.

These statistics are not ideal for our purposes for several reasons: first, thinking of reported field sizes as shown in the Oil and Gas Journal for, say, Oklahoma in 1954 as a sample from some true underlying size distribution of oil fields, the sample is truncated. That is, of all fields discovered as of a given year, only those "fields with production of at least 1000 barrels per calendar day"^{††} are included in the sample. Second, there is a great deal of reporting bias present in these estimates of field sizes, especially in estimates of the size of "younger" fields before substantial production experience with them has accrued. Third, the meaning of the definition of field size as "total ultimate primary reserves"^{†††}

[†]The sources of these statistics are the Oil and Gas Journal Annual Review Issues for 1952, 1954, 1956, 1960, and 1964. See Table I.

^{††}cf. Oil and Gas Journal, Annual Review Issue, Vol. 58, No. 4 (January 25, 1960) p. 163. Major fields with an ultimate recovery estimate of 100 million barrels or more are included regardless of daily production.

^{†††}Ibid.

recoverable from the field changes as the technology of recovery changes. Fourth, the usual definition of a field as consisting of one or more pools vertically separated from one another but having similar areal boundaries or outline is flexible and is not operationally precise in every instance; e.g. the Davenport "field" in Oklahoma is listed as one "field" in the Oil and Gas Journal, but reported in terms of six subdivisions (fields) in the International Oil and Gas Development Yearbook.[†]

However, we can use these statistics as an illustrative vehicle, keeping in mind that any company undertaking a study of field size distributions for the purposes mentioned earlier can improve the quality of such data by confining their analysis to geological basins instead of geographical areas such as states, and by using more refined sources of information.^{††}

Table I lists the issues from which the data used here was taken, as well as the definitions of reserves used in each year and the criteria for including a field in the list.

2.2 Methodology

For both Oklahoma and South Louisiana we carried out these steps:

1. For each of the years 1952, 1954, 1956, 1958, 1960, 1962, and 1963 we calculated an estimate of ultimate primary reserves of each field by adding "proved remaining reserves"

[†]1962, Part II, Vol. XXXII, p. 137 ff.

^{††}See [3] for example.

to "cumulative production to date", and then ordered these estimates from smallest to largest.

2. The ordered estimates were plotted as described in subsection 2.3 and displayed in Graphs 1 through 9.
3. A Lognormal distribution was fitted to each plot and its parameters were estimated two different ways. (Table III).
4. Since the shape of a Lognormal distribution is completely determined once its two parameters are specified, we summarized the character of each of the nine graphs by specifying the parameters of the Lognormal distribution fitted to it. (Graph 9).
5. In Graph 9 estimates of the mean, median, and mode of each frequency histogram of field sizes are displayed so as to trace the manner in which they change over time.

As stated earlier, the probability distribution of field sizes remaining to be discovered in an area as of a given point in time is of critical importance in assessing the expected profitability and the "risk" of pursuing an exploration program in that area at that point in time. The effect of timing on the expected profitability and "riskiness" of an exploration program can only be explicitly assessed if one has a systematic way of showing how the odds of discovering economically viable fields changes with time. To this end we carried out steps 6 and 7.

6. We used the parameter estimates shown in Table III together with some hypothetical subjective probability judgements about the true underlying size distribution of oil fields in barrels to compute a probability distribution of field sizes remaining to be discovered in South Louisiana as of 1952 and as of 1964.
7. We calculated the mean and variance of field sizes remaining to be discovered in South Louisiana as of 1952 and as of 1964 using the results of step 6 as an example of how the first two moments of the probability distribution of field sizes remaining to be discovered changes with time.

TABLE I

SOURCES OF DATE USED IN EXHIBITS 1 THROUGH 9⁺

<u>Year</u>	<u>Vol. No. and Page</u>	<u>Definition of Reserves</u>	<u>Criteria for Inclusion</u>
1952	Vol. 51 No. 38 p. 289	"...estimates of proved remaining reserves of crude oil, condensate, and cycled products for the country's larger producing fields."	"Larger producing fields"
1954	Vol. 53 No. 39 p. 197	same as 1952	same as 1952.
1956	Vol. 55 No. 4 p. 159	same as 1952	"Those major fields with an estimated ultimate recovery of 100 million barrels or more are... included here regardless of present daily production. Fields included all produce 1,000 barrels or more per calendar day."
1958	Vol. 57 No. 4 p. 141	"...estimated remaining reserves of crude oil and condensate for the larger fields of the United States. Eligibility for this list requires a 1958 production rate of at least 1,000 barrels per calendar day. Figures refer to primary recovery only."	same as 1956
1960	Vol. 59 No. 5 p. 126	The last sentence above is amended to read: "Figures...in the majority of cases refer to primary recovery only."	same as 1956
1962	Vol. 61 No. 4 p. 172	same as 1960	same as 1956

⁺
All Vol. references are to the Oil and Gas Journal.

2.3 Plotting the Data and Curve Fitting

The total number N of fields in, say, South Louisiana discovered to date constitutes a sample of fields from the true underlying size distribution of fields in South Louisiana, and the number $n < N$ of fields reported in the Oil and Gas Journal constitute a very particular sub-sample from this sample.

Provided that we know N , we may use this sub-sample to test the Lognormality of the size distribution of oil fields in South Louisiana as follows:

1. List the n sub-sample observations in order of size, from smallest to largest.
2. Consider the k th largest observation as an estimate of the $(N-n+k)/(N+1)$ st fractile of the true underlying distribution of field sizes.[†]
3. Plot the fractile estimates on Lognormal probability paper.
4. Fit a straight line to the data.

If the sample observations are from a Lognormal population, then the plotted points should lie approximately on a straight line. The larger the number of sample observations and the closer the fit to a straight line, the more reasonable the assumption of Lognormality becomes.

In order to determine N , we first counted the number of fields in Oklahoma and in South Louisiana as listed in the International Oil and Gas Development Yearbook 1962.^{††} In some instances, the definition of a field in this reference differed from that given in the Oil and Gas Journal. The fields listed in the former reference were aggregated when necessary so

[†]Any value which is both (a) equal to or greater than a fraction $.f$ of the values in the set and (b) equal to or less than a fraction $(1-.f)$ of the values in the set is a $.f$ fractile of the set.

^{††}Part II, Vol. XXXII, p. 137ff.

as to make the field definition correspond with that being used in the Oil and Gas Journal and then N was calculated. This aggregation had a significant effect on N only in Oklahoma, where many of the fields listed in the Oil and Gas Journal are listed as two or more geographic subdivisions; e.g. Davenport in the Oil and Gas Journal is listed in the Yearbook as Davenport, Davenport North, Davenport Northeast, Davenport South, Davenport Southeast, and Davenport West.

TABLE II

	<u>Number of Fields Discovered</u>	
	<u>As Listed in Yearbook</u>	<u>Aggregated</u>
North Louisiana	328	271
South Louisiana--offshore	131	125
--onshore	641	508
--combined	772	633
Oklahoma	2704	1361

Here, however, another difficulty arises, for N is well over 100 for both Oklahoma and South Louisiana. The largest sample observation in Oklahoma corresponds to a .9996 fractile--and Lognormal probability paper presently in use allows one to plot fractile numbers of the order of .99 or less. Since we are particularly interested in the behavior of the right tail of the distribution, ordinary Lognormal probability paper is unsuitable.

One way of overcoming this difficulty is to plot sample values against standardized Normal units as shown in Exhibits 1 through 9. The procedure used here is justified mathematically and described in detail in Appendix A.

As with Lognormal probability paper, if the sample observations are from a lognormal population then the plotted points in the upper right tail of the plot should lie approximately on a straight line.

Once the data is plotted, a straight line can be fitted by eye,[†] and the resulting line used to give graphical estimates μ and σ of the parameters μ and σ of a Lognormal distribution as described in Appendix A. Alternately, one may estimate μ and σ using actual data points. Both ways of estimating the parameters of Lognormal distributions fitted to the plots of Exhibits 1 through 9 were used and the results are displayed in Table III.

A more informative display of properties of these fitted distributions--means, medians, and modes--is given in Exhibit 9.

2.4 Discussion of Exhibits 1 through 9

Some important attributes of these Exhibits are:

1. All of the plots, instead of being linear, have a pronounced downward bend in the left tail of plotted points.
2. By 1963, a straight line fits the right tail of both the Oklahoma and South Louisiana histograms reasonably well.
3. The estimated mean of reported field sizes for South Louisiana steadily increases from 1952 to 1963, while that of Oklahoma slightly increases from 1952 to 1960 and then decreases. By 1963, the estimated mean

[†]For a discussion of errors introduced by fitting a line by eye see [4].

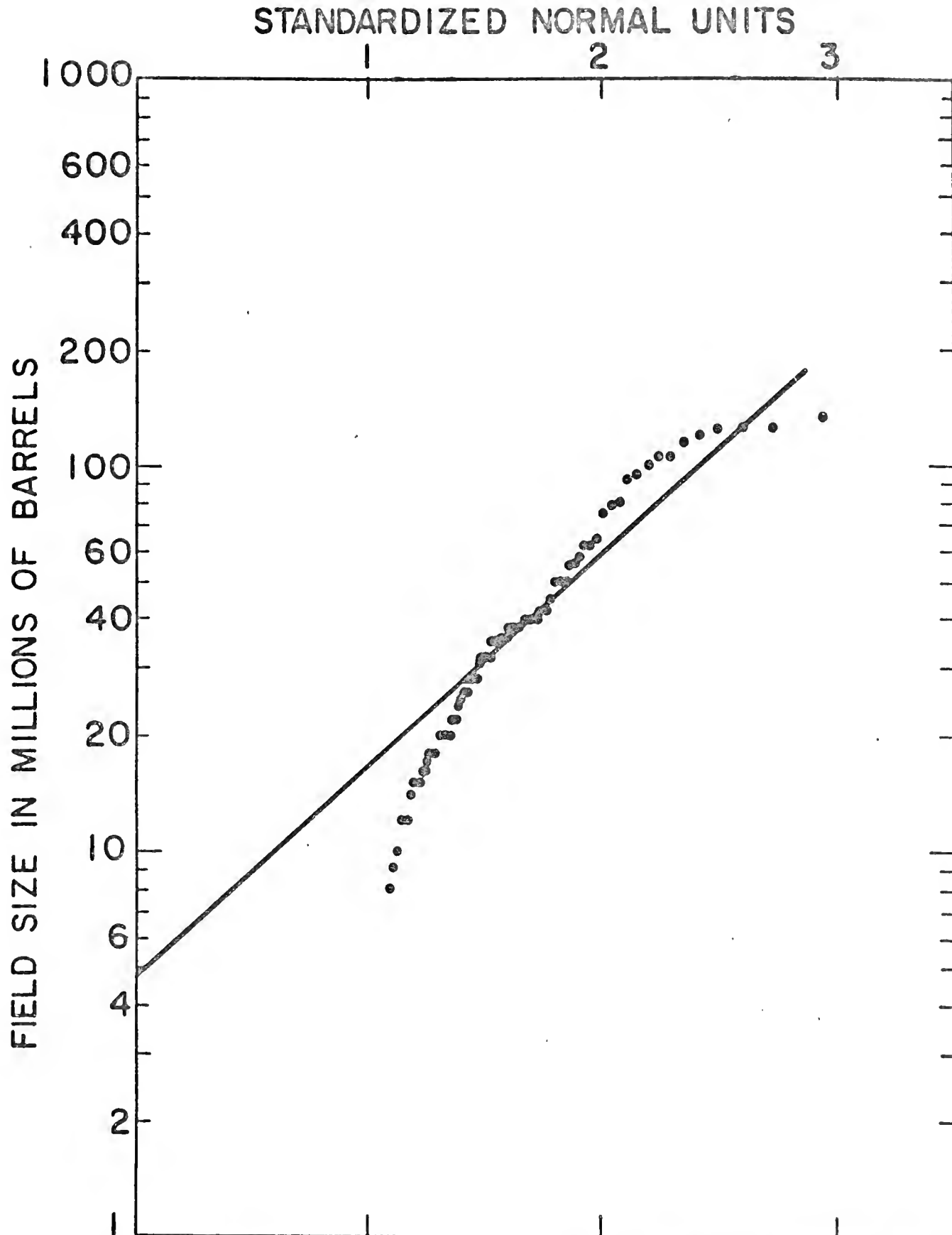


Exhibit 1 - FRACTILES OF ULTIMATE PRIMARY RESERVES IN SOUTH LOUISIANA - OFFSHORE AND ONSHORE - AS OF 31 DECEMBER 1952 (N=633, n=86)

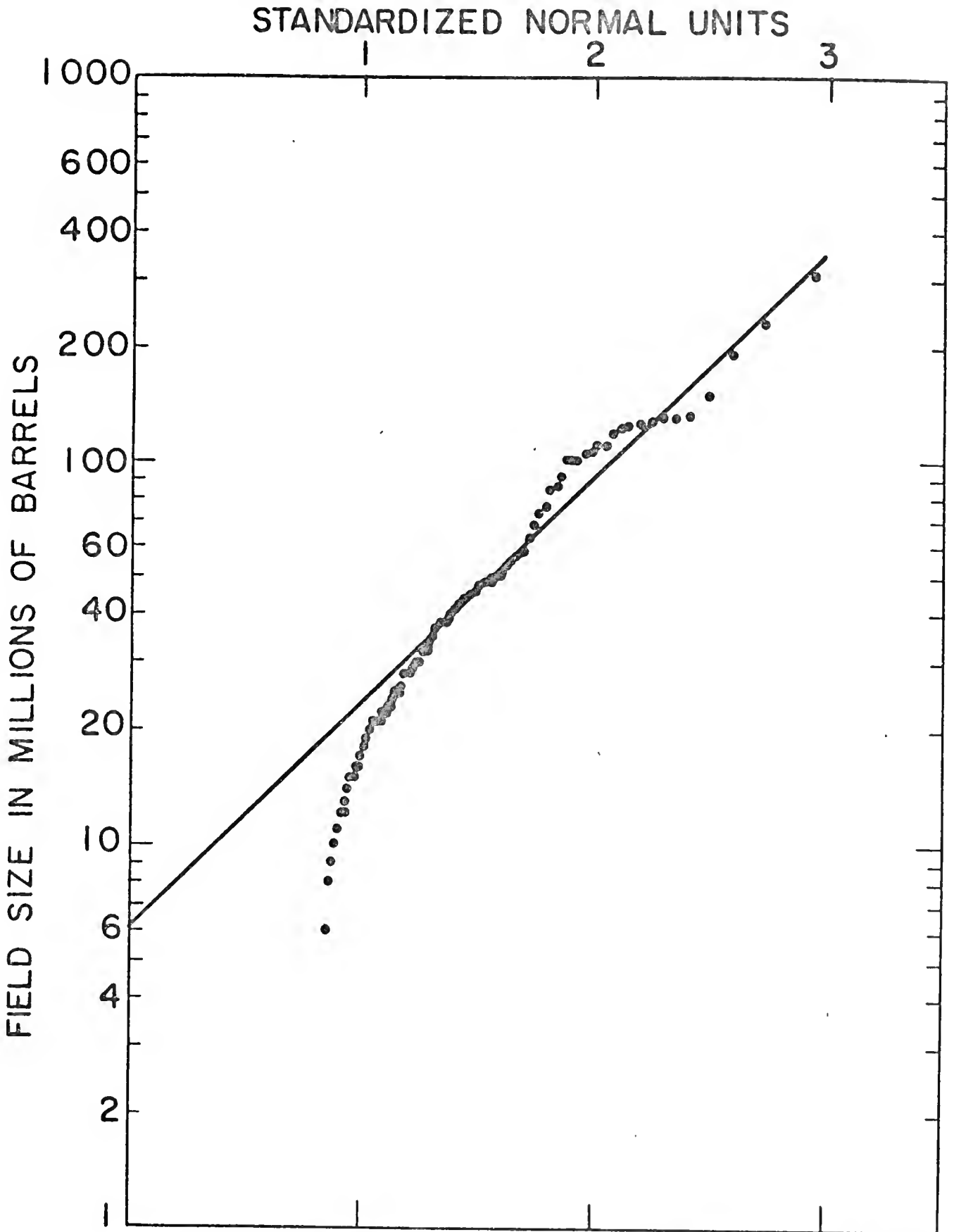


Exhibit 2 - FRACTILES OF ULTIMATE PRIMARY RESERVES IN SOUTH LOUISIANA - OFFSHORE AND ONSHORE - AS OF 31 DECEMBER 1956 (N=633, n=124)

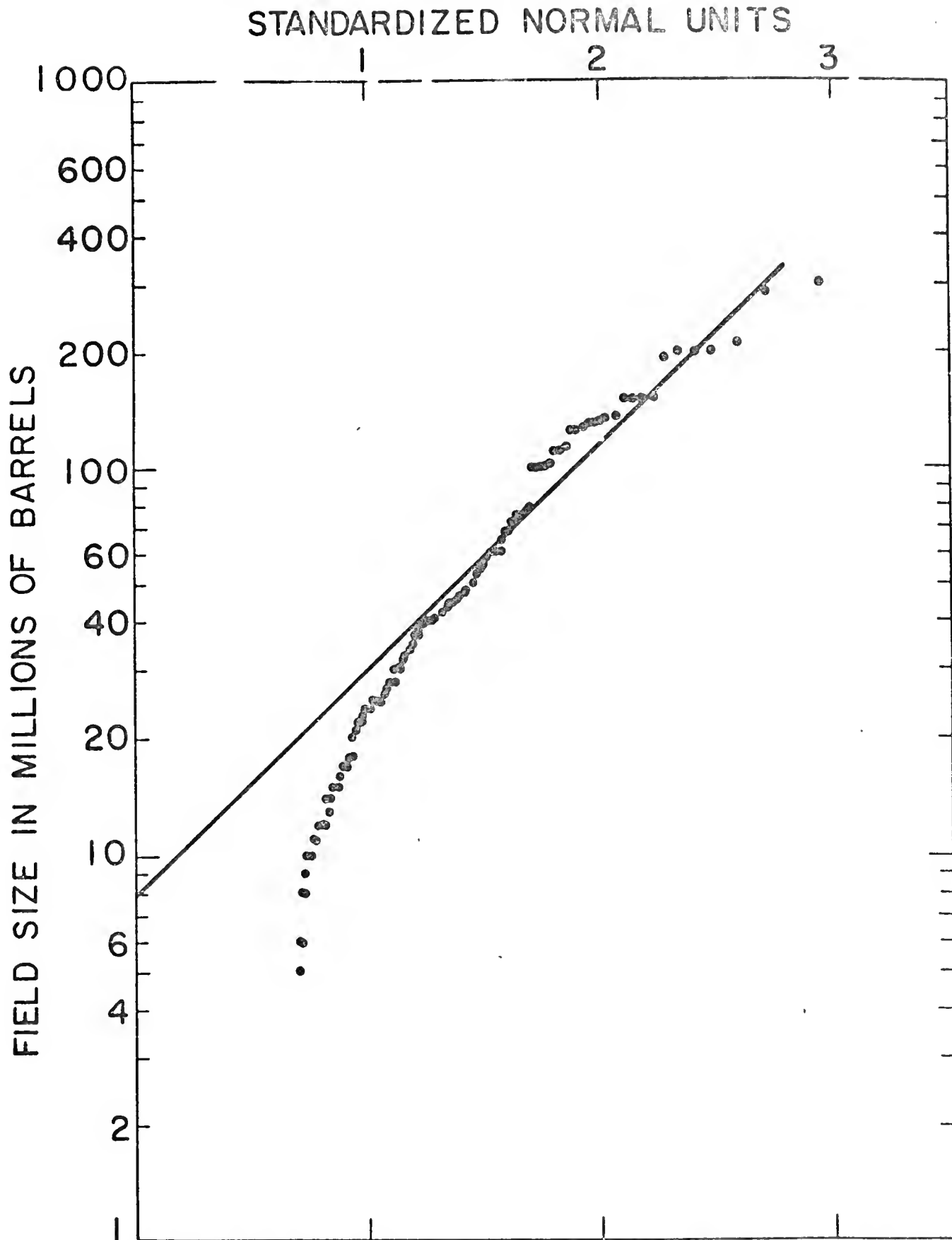


Exhibit 3-FRACTILES OF ULTIMATE PRIMARY
RESERVES IN SOUTH LOUISIANA -
OFFSHORE AND ONSHORE - AS OF
31 DECEMBER 1960 (N=633, n=152)

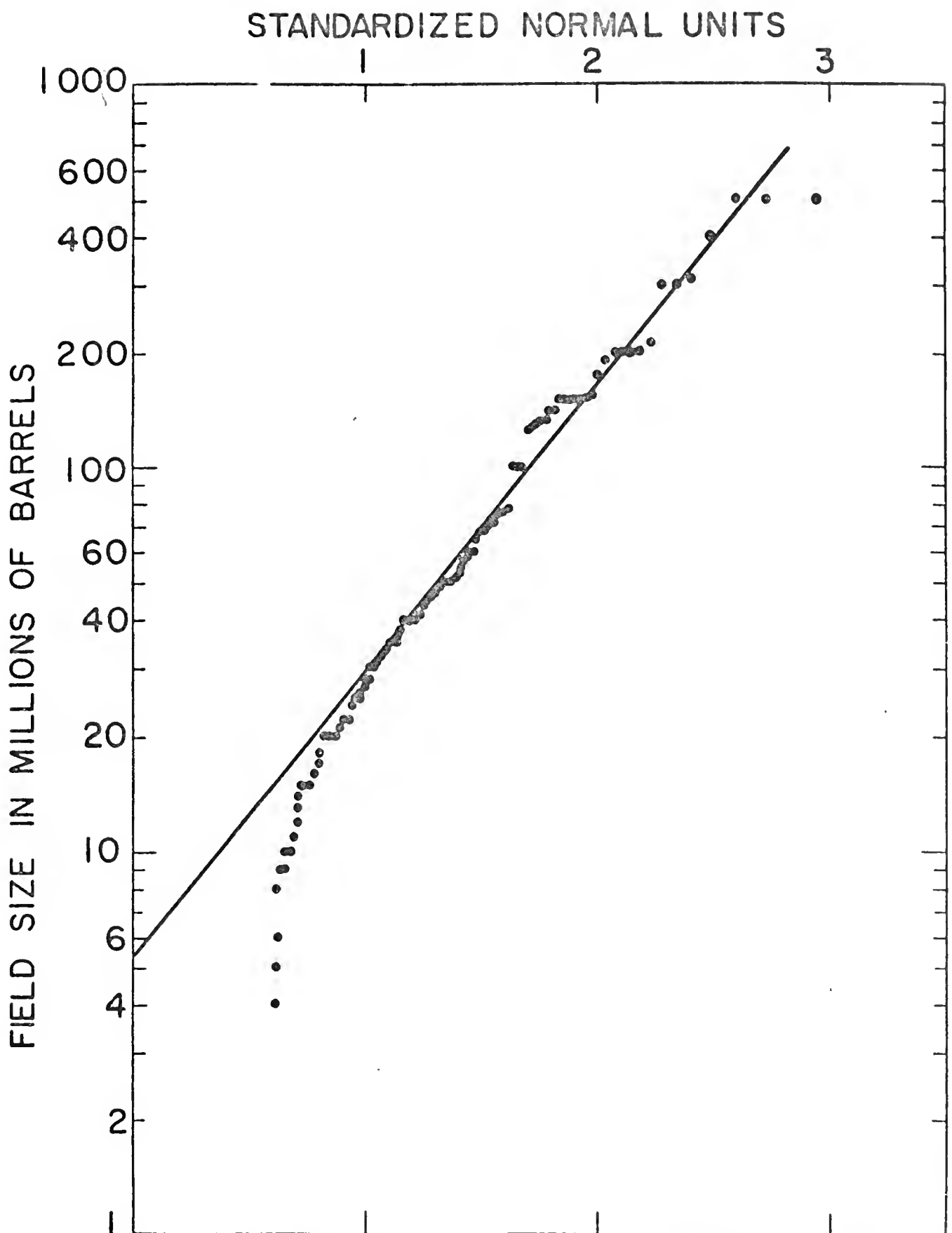


Exhibit 4 - FRACTILES OF ULTIMATE PRIMARY
RESERVES IN SOUTH LOUISIANA -
OFFSHORE AND ONSHORE - AS OF
31 DECEMBER 1963 (N=633, n=171)

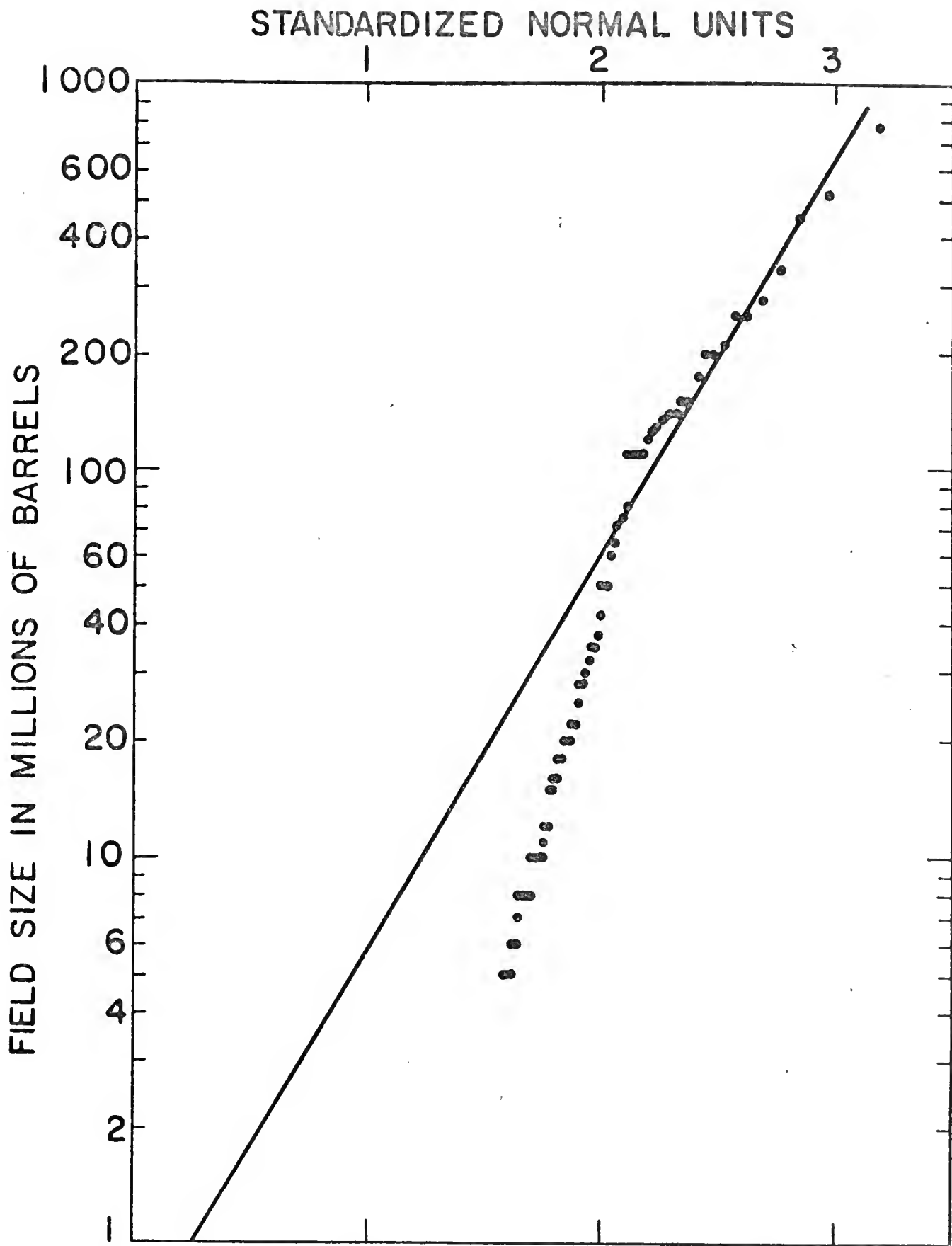


Exhibit 5 - FRACTILES OF ULTIMATE PRIMARY
RESERVES IN OKLAHOMA - AS OF
31 DECEMBER 1952 (N=1361, n=75)

STANDARDIZED NORMAL UNITS

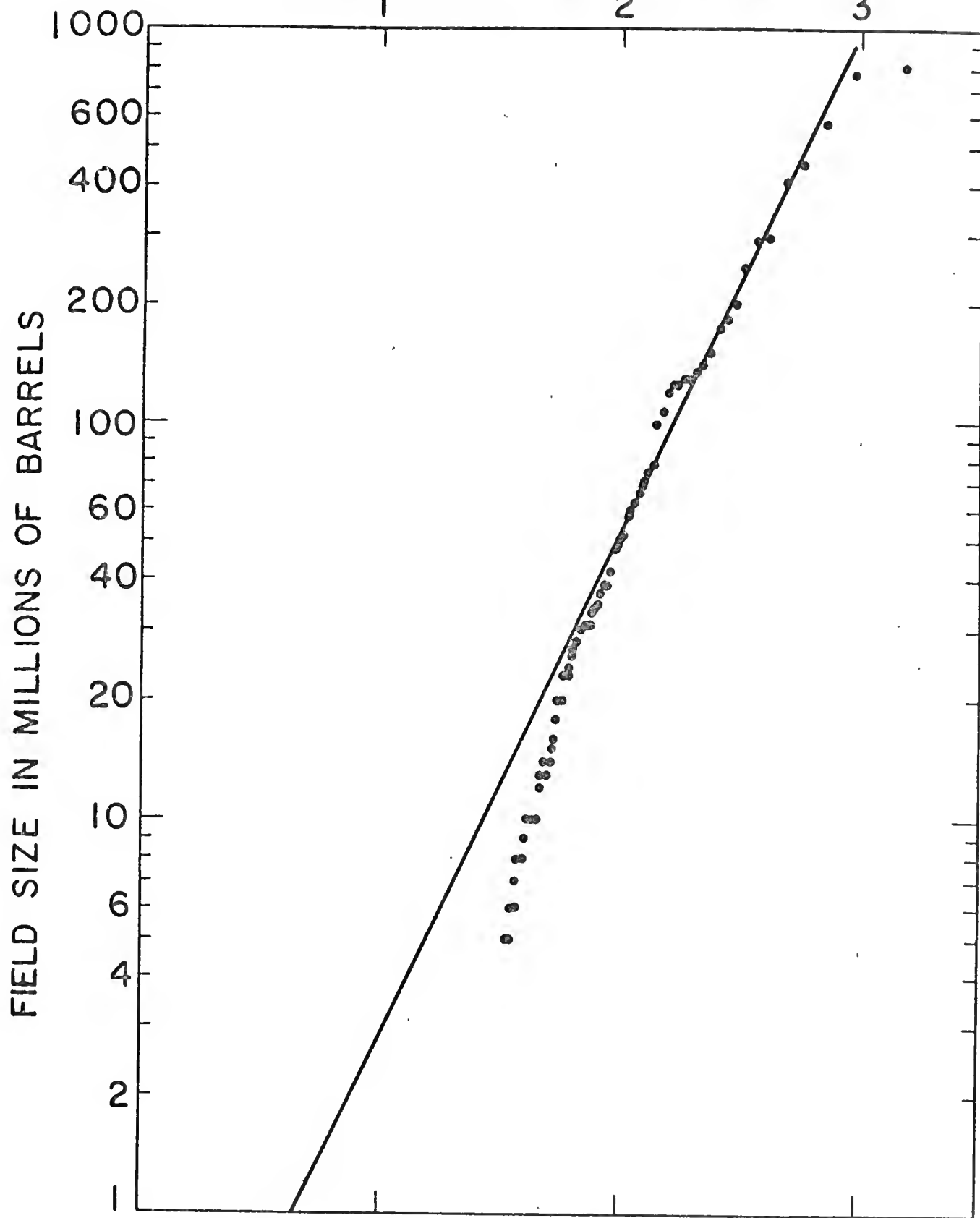


Exhibit 6 - FRACTILES OF ULTIMATE PRIMARY RESERVES IN OKLAHOMA - AS OF 31 DECEMBER, 1956 (N=1361, n=85)

STANDARDIZED NORMAL UNITS

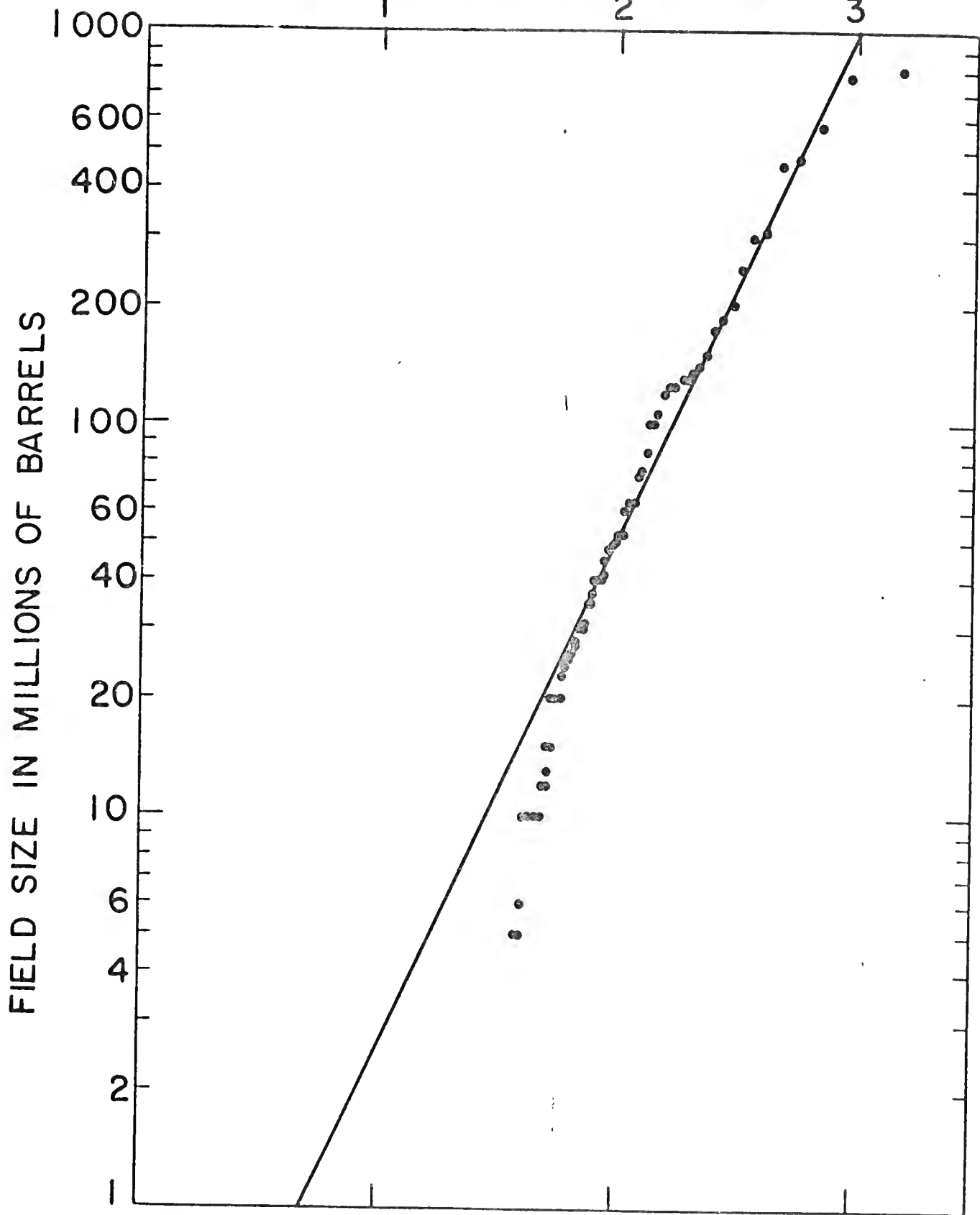


Exhibit 7-FRACTILES OF ULTIMATE PRIMARY
RESERVES IN OKLAHOMA - AS OF
31 DECEMBER 1960 (N=1361, n=75)

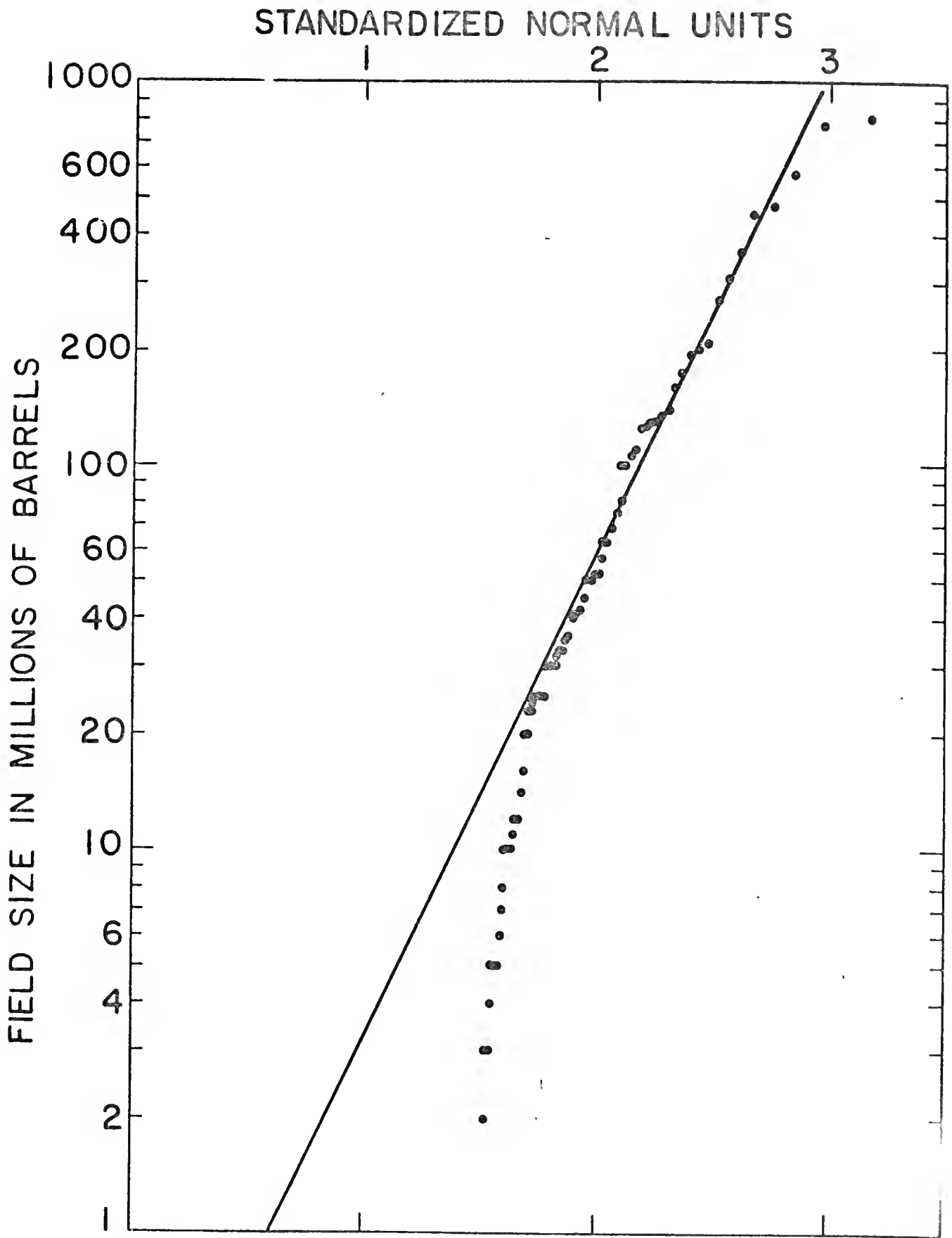
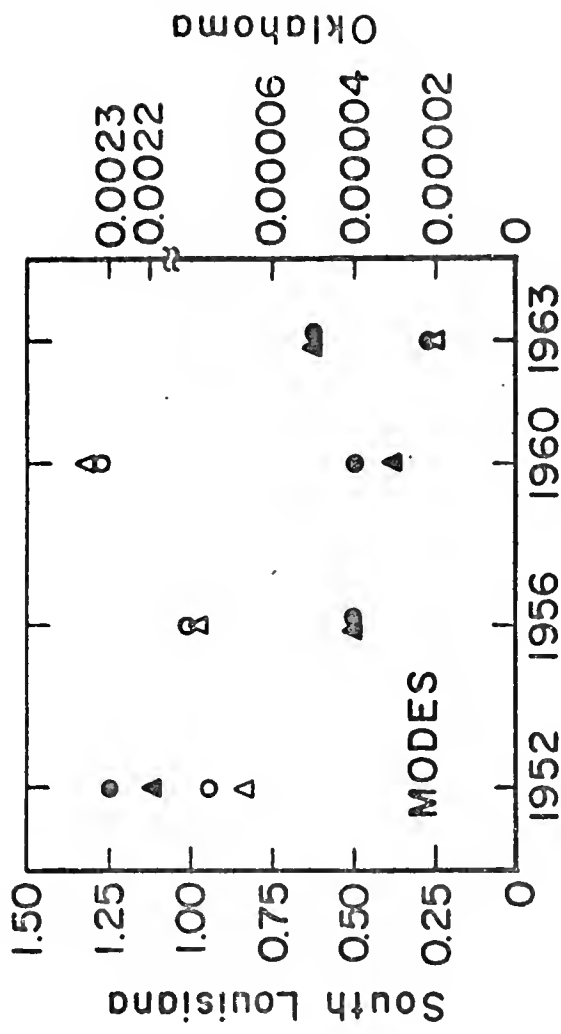
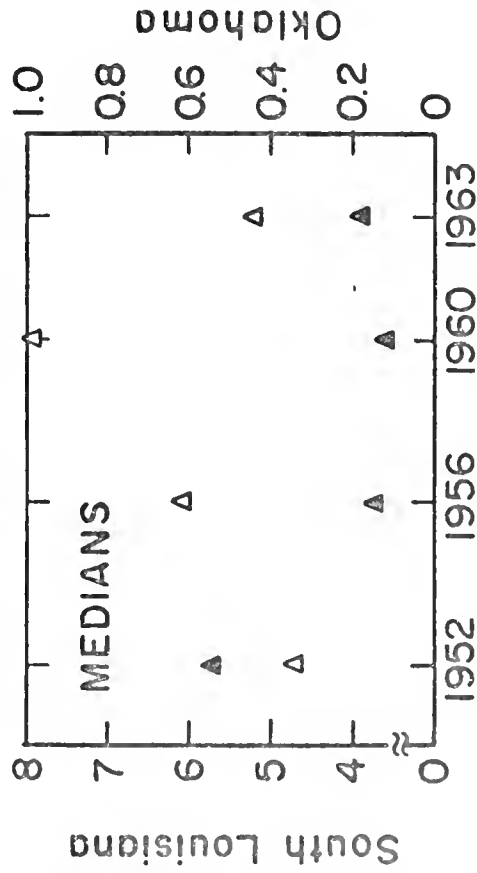
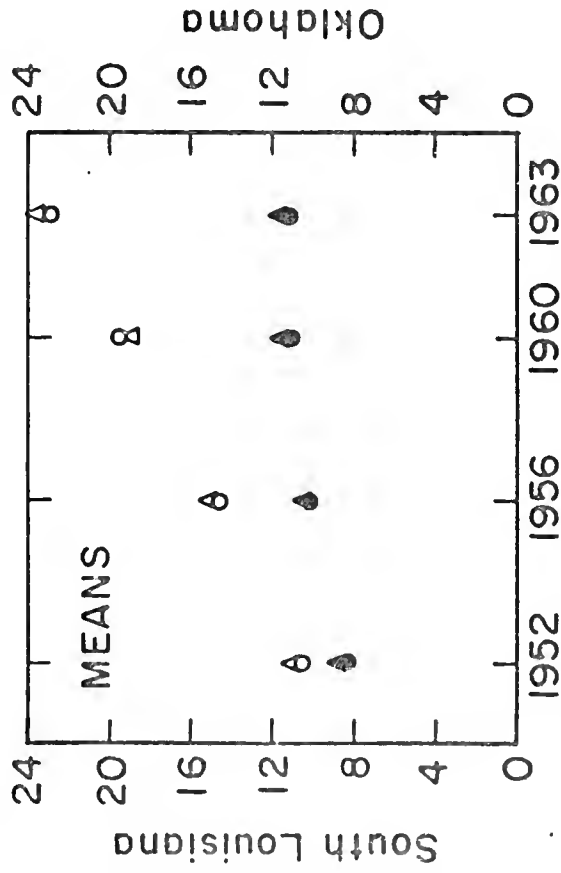


Exhibit 8 - FRACTILES OF ULTIMATE PRIMARY
RESERVES IN OKLAHOMA - AS OF
31 DECEMBER 1963 (N=1361, n=86)

ESTIMATES OF STATISTICS OF SIZE DISTRIBUTIONS FOR OKLAHOMA AND SOUTH LOUISIANA



- South Louisiana calculated
- △ South Louisiana graphical
- Oklahoma calculated
- ▲ Oklahoma graphical

reported field size in South Louisiana is roughly twice that of Oklahoma.

4. Estimates of median and modal field sizes are extremely small by comparison with estimated means so that the distributions are all highly skewed to the right.

The downward bend in the left tail of all points is probably due in great part to the peculiar criteria for inclusion of fields in the Oil and Gas Journal listing--1000 or more barrels of production per calendar day. This undoubtedly results in many "older" fields with ultimate primary reserves of 1 to 20 million barrels being omitted from the listing, and resultant the "thinness" of listed observations in this range causes the downward bend. (A careful cataloguing of smaller fields would determine whether or not this conjecture is true.)

If we focus on the right tail of the plot, however, and discount the smaller values by fitting a straight line to those fields of a size greater than 40 million barrels in South Louisiana and in Oklahoma then the fit looks reasonable. As can be seen, this was in fact what was done in the visual fitting of these lines. The justification for fitting only plotted points far out in the right tail with a straight line and then using the line to make inferences about the parameters of the whole distribution is that

IF THE TRUE UNDERLYING DISTRIBUTION IS LOGNORMAL, AS THE NUMBER OF "LARGE" SAMPLE OBSERVATIONS INCREASES, A STRAIGHT LINE FITTED TO PLOTS OF SAMPLE POINTS IN THE RIGHT TAIL OF HISTOGRAMS LIKE THOSE DISPLAYED IN EXHIBITS 1 THROUGH 9 WILL ASYMPTOTICALLY LEAD TO ESTIMATES OF μ AND σ CLOSE TO THEIR TRUE VALUES WITH HIGH PROBABILITY.

Notice that both graphical and calculated estimates of parameters and of means, medians and modes are extremely close. (Exhibit 9). The regularity

TABLE III
ESTIMATES OF PARAMETERS OF SIZE DISTRIBUTIONS (a,b,c)

	(1) $\hat{\mu}$	(2) $\hat{\sigma}$	(3) $\hat{\sigma}_o$	(4) \hat{h}	(5) \hat{h}_o	(6) MEAN	(7) MEAN _o	(8) MEDIAN	(9) MODE	(10) MODE _o
South Louisiana	1952	1.31	1.27	.578	.622	11.25	10.59	4.75	.84	.95
	1956	1.81	1.34	.550	.559	15.18	14.88	6.10	.99	1.02
	1960	2.07	1.35	.562	.549	19.20	19.63	7.90	1.33	1.28
	1963	1.65	1.72	.332	.337	23.44	22.87	5.20	.26	.27
Oklahoma	1952	.60	2.34	.180	.183	8.82	1.49	.55	.0022	.0023
	1956	-1.90	2.93	.117	.117	10.85	10.60	.15	.00004	.00004
	1960	-2.08	3.02	.110	.111	11.82	11.13	.125	.00003	.00004
	1963	-1.74	2.91	.118	.120	11.96	11.30	.175	.00005	.00005

(a) Estimates in columns (1), (2), (4), (6), (8) and (9) are calculated directly from the straight line approximations in Exhibits 1 through 9.

(b) Estimates in columns (3), (5), (7) and (10) are calculated using $\hat{\mu}$ and 10 actual data points.

(c) Columns (6) through (10) are in millions of barrels.

of behavior of estimates of means seems to tag them as better indicators of the change of the histograms with time.

3. The Probability Distribution of Field Sizes Remaining to be Discovered

In order to calculate the probability distribution of field sizes remaining to be discovered in an area at a given point in time we must make some assumptions about the true size distribution of fields in the area, total ultimate recovery from the area, and the manner in which the histogram of reported field sizes and the true size distribution of fields are interrelated.

3.1 Definitions and Assumptions

For a given area and time periods $t=0,1,2,\dots$, define

r_t - reported size in barrels of a generic field discovered by time t ,

s - true size in barrels of a generic field,

z - size of a generic field undiscovered by t (unadjusted for reporting bias),

R_t - total ultimate primary reserves recoverable from fields discovered by time t ,

S - total ultimate primary reserves recoverable from all fields in the area.

We make three assumptions:

I - The random variable \tilde{s} is Lognormally distributed with parameter (μ_s, σ_s^{-2}) :

$$\tilde{s} \sim f_L(s | \mu_s, \sigma_s^{-2}) .$$

II - The random variable \tilde{r}_t is Lognormally distributed with parameter (μ_t, σ_t^{-2}) :

$$\tilde{r}_t \sim f_L(r_t | \mu_t, \sigma_t^{-2}) .$$

3.2 Derivation

We may informally regard the process of discovering fields as hypergeometric--like sampling from a Lognormal population in which "area" under the probability distribution function of \tilde{s} is "used up" as sampling progresses; i.e. as more and more fields are discovered. In [1][†] it is shown that as the number of discovered fields increases, the frequency histogram of these fields asymptotically approaches a Lognormal functional form if I above is true.

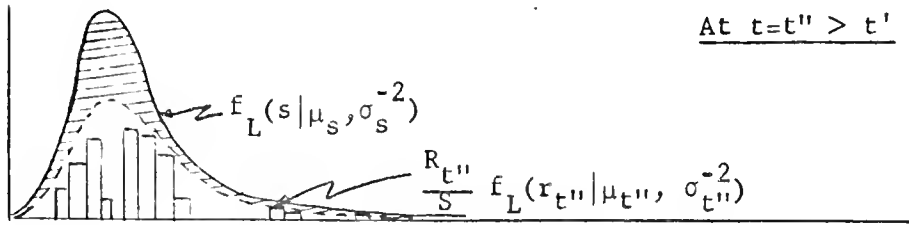
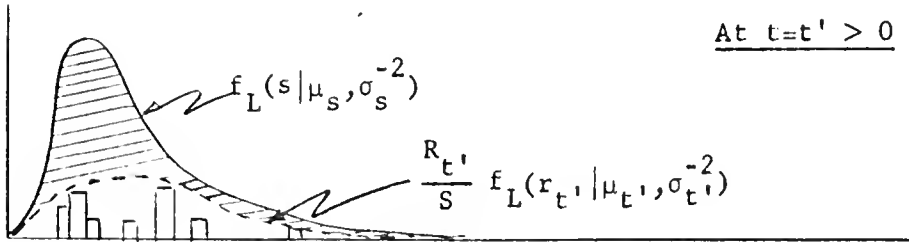
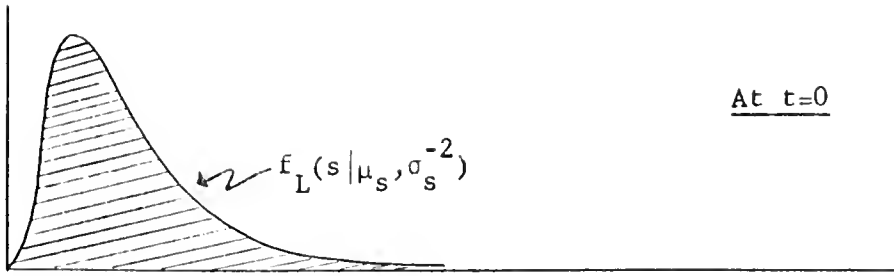
More precisely, I and II imply that given R_t , S , (μ_t, σ_t^{-2}) , and (μ_s, σ_s^{-2}) , the probability distribution of field sizes remaining to be discovered \tilde{z} is

$$\kappa [f_L(z|\mu_s, \sigma_s^{-2}) - \frac{R_t}{S} f_L(z|\mu_t, \sigma_t^{-2})] ,$$
$$z > 0, \sigma_t > 0, \sigma_s > 0 ,$$
$$-\infty < \mu_t < +\infty ,$$
$$-\infty < \mu_s < +\infty ,$$

where κ is a normalizing constant.

The idea of characterizing the probability density function of \tilde{z} as in (1) may be shown graphically this way:[†]

[†]This view of the process is simply a probabilistic adaptation of a deterministic model proposed by J.J. Arps and T.G. Roberts, [3].



The probability density of field sizes remaining to be discovered is roughly proportional at each point in time to the height of the shaded area. While we could in fact calculate this density function more accurately, a great deal of analytical simplicity is achieved by first approximating the shape of the sample histogram with a Lognormal distribution and then using (1), as we shall see. One can use the methods described in section 2 to do this. And R_t may be directly calculated.

Here we assume that the parameters μ_s , and σ_s^{-2} have been derived using the subjective judgements of geologists in the manner described in Chapter 6 of [1], and that S has been estimated by a gross volumetric analysis of the area.

Using (1), we may directly calculate moments of the distribution of \tilde{z} .

Its mean and variance are

$$E(\tilde{z}) = \kappa e^{\mu_s + \frac{1}{2}\sigma_s^2} - \kappa \frac{R_t}{S} e^{\mu_t + \frac{1}{2}\sigma_t^2},$$

$$V(\tilde{z}) = E(\tilde{z}^2) - E^2(\tilde{z}) = \kappa e^{2\mu_s + 2\sigma_s^2} - \kappa \frac{R_t}{S} e^{2\mu_t + 2\sigma_t^2} - E^2(\tilde{z}).$$

and the kth partial moment of \tilde{z} about the origin is as shown in formula (2) of Appendix B.

3.3 An Hypothetical Example

To illustrate how these results may be used, suppose we wish to determine the probability distribution of \tilde{z} for South Louisiana as of 1963.

A geological and geophysical evaluation has led us to specify that as of 1963,

$$\mu_s = 1.5, \quad \sigma_s^2 = 3.0, \quad S = 25000 \times 10^6.$$

Letting 1900 be $t=0$, taking estimates of μ_{63} , and σ_{63}^2 from Table III, and calculating R_{63} , we may use (1) to calculate the probability density function $f_z(\cdot)$ of \tilde{z} as of 1963. Aside from the normalizing constant κ_{63} , some right tail values of $f_z(\cdot)$ for 1963 are shown in column (6) of Table V.

In a similar fashion we calculated some right tail values of $\kappa_{52}^{-1} f_z(\cdot)$ for comparative purposes. These are shown in column (5) of Table IV.

Values such as those displayed in Table IV were used to plot the curves of Exhibits 10 and 11. Note that these exhibits give a typical example of how the right tail of the distribution of \tilde{s} is "used up" as the 86 fields used to calculate fractile estimates as of 1952 (Exhibit 1) increase to 171 fields in 1963. (Exhibit 4)

Using the formulae of Appendix B, we may calculate the moments shown in Table VI.

TABLE IV

PROBABILITY DENSITY FUNCTIONS OF FIELD SIZES \tilde{z} REMAINING TO BE DISCOVERED

(in 10^6 bbls.)	$f_L(z \mu_s, \sigma_s^2)$	$\frac{R_{52}}{S} f_L(z \mu_{63}, \sigma_{63}^2)$	$\kappa_{52}^{-1} f_z(\cdot)$	$\frac{R_{63}}{S} f_L(z \mu_{63}, \sigma_{63}^2)$	$\kappa_{63}^{-1} f_z(\cdot)$
4	.361	.047	.313	.178	.183
5	.289	.038	.251	.145	.144
6	.238	.032	.206	.120	.118
7.5	.185	.025	.160	.095	.090
10	.130	.018	.112	.067	.063
20	.0498	.0068	.0430	.0270	.0228
30	.0250	.0036	.0214	.0145	.0105
50	.0110	.0015	.0095	.0062	.0048
75	.0052	.0007	.0045	.0030	.0022
100	.0029	.0004	.0025	.0017	.0012
200	.00064	.00009	.00055	.00040	.00024
300	.00026	.00004	.00022	.00015	.00011
400	.00013	.00002	.00011	.00008	.00005
500	.00009	.00001	.00008	.00004	.00005
1000	.000014	.000002	.000012	.000007	.000007

TABLE V

	<u>P($\tilde{z} > z$)</u>	
<u>z</u>	<u>1952</u>	<u>1963</u>
4	.525	.493
5	.475	.461
6	.435	.398
7.5	.388	.355
10	.327	.294
20	.198	.166
30	.142	.116
40	.110	.089
50	.087	.067
75	.052	.036
100	.040	.030
200	.015	.011
300	.009	.006
400	.0055	.004
500	.0033	.002
1000	.0010	.0006

	<u>1952</u>	<u>1963</u>
Mean of \tilde{z}	21.04	16.84
Variance of \tilde{z}	8486.9	4984.6
Partial expectation of $\tilde{z} > 40 \times 10^6$ bbls.	15.2	10.8
Conditional Expected Field Size given $\tilde{z} > 40$	138.2	121.3

It follows, that as of 1963, under the assumptions made earlier in this example about \tilde{s} and \tilde{r}_t that:

- (1) The probability of a newly discovered field being larger than 40 million barrels is .089.
- (2) Conditional upon discovering a new field of 40 million barrels or larger, its expected (reported) size is 121.8 million barrels.

4. Summary

We have shown that the Lognormal functional form is a reasonable and useful tool to use in the course of analytically characterizing the probability distribution of the size of fields remaining to be discovered in a basin. In section 2 we developed a method for fitting Lognormal functional forms to empirical histograms of reported field sizes even when the samples from which these histograms are constructed are truncated. We then showed how to calculate estimates of the parameters of the Lognormal distributions according to which the samples were assumed to have been generated.

Finally, in section 3 we illustrated by example how the results of the analysis of sections 1 and 2 can be used to calculate a probability distribution of ~~fieldsizes~~ remaining to be discovered in an area, as well as to calculate the moments of this distribution.

EXHIBIT 10

RIGHT TAIL OF PROBABILITY

DISTRIBUTIONS OF \tilde{r}_{52} , \tilde{r}_{63} , AND \tilde{s}

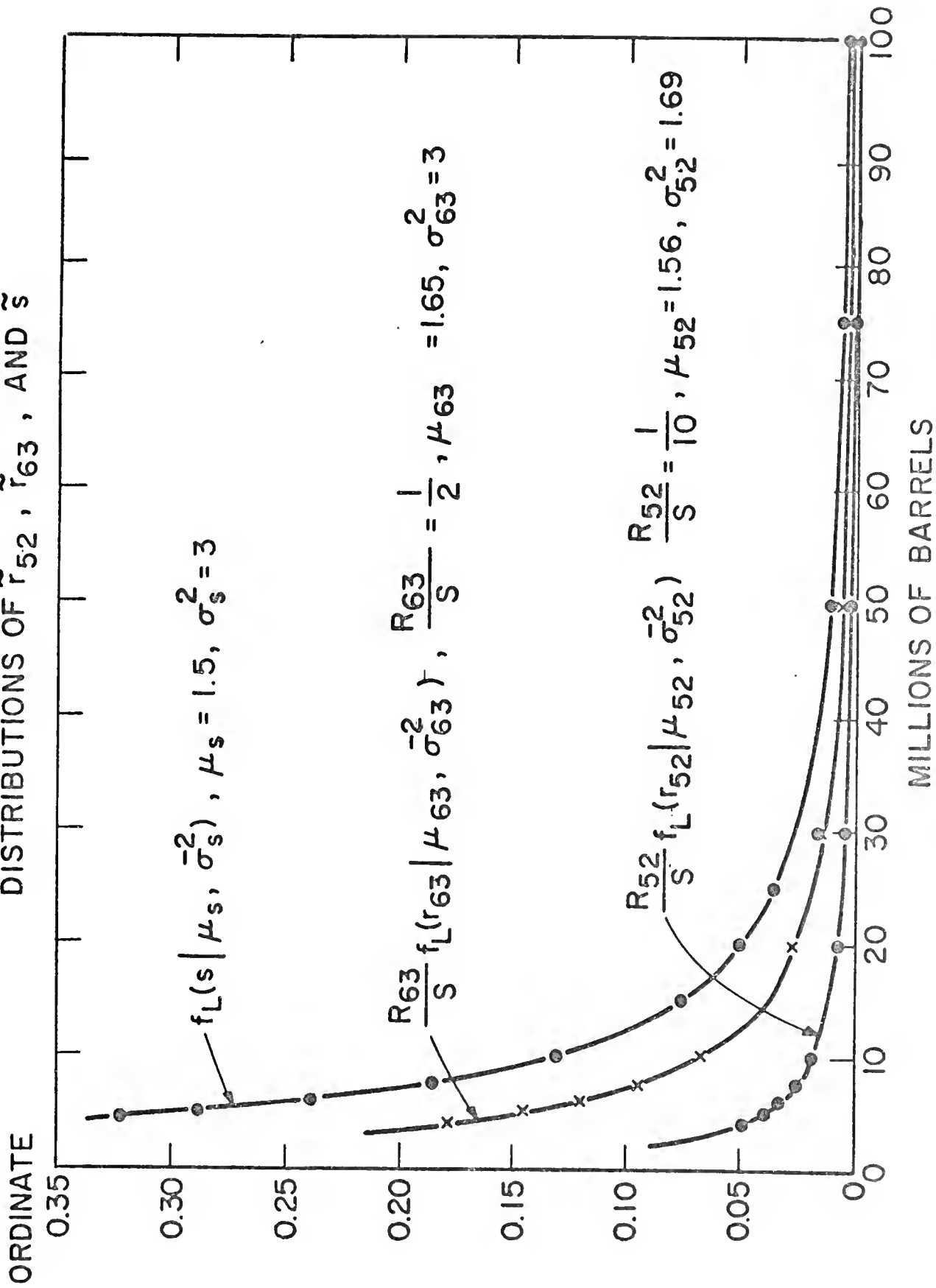
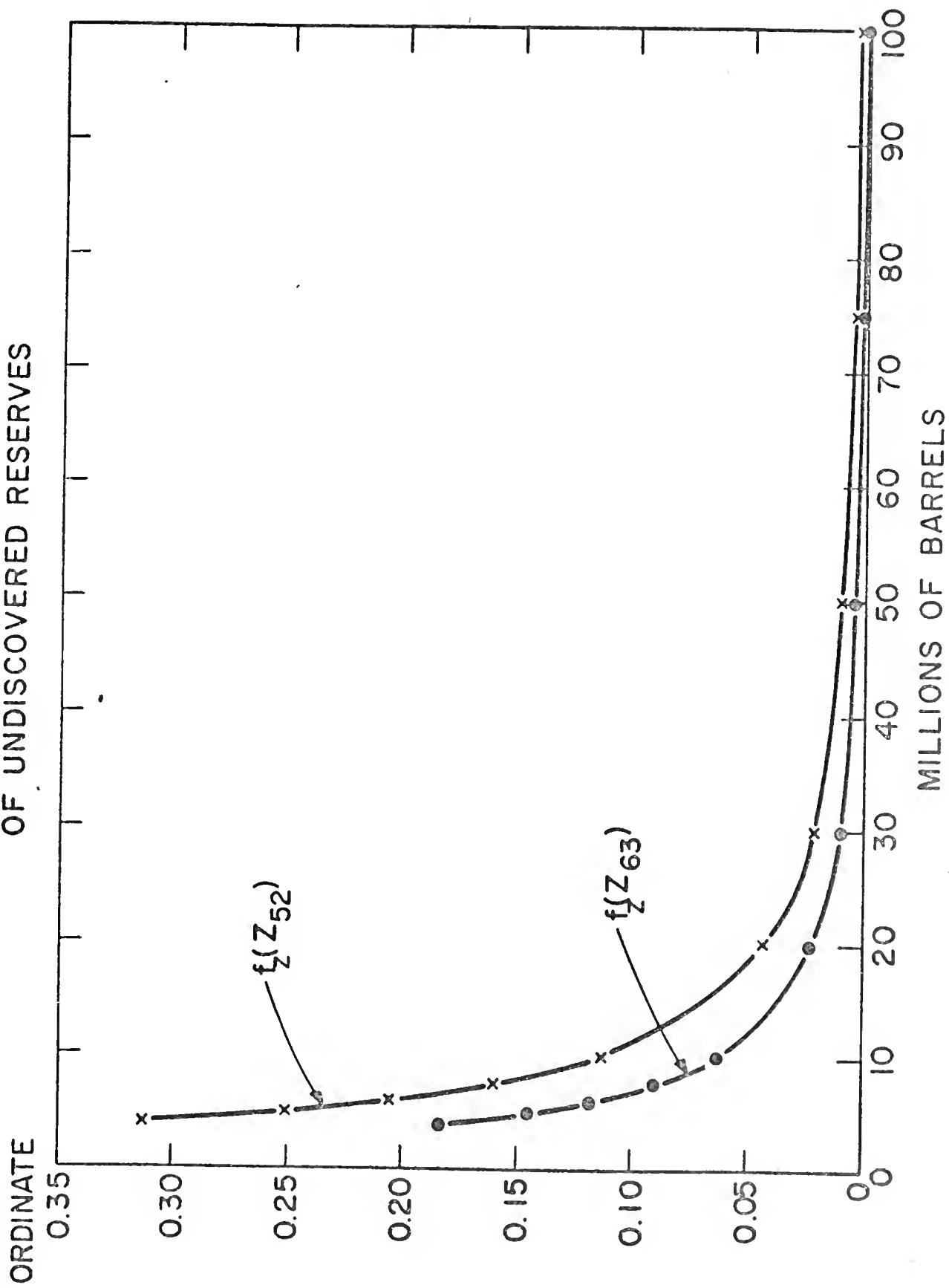


EXHIBIT II

RIGHT TAIL OF PROBABILITY DISTRIBUTION OF SIZE
OF UNDISCOVERED RESERVES



REFERENCES

- [1] G. M. Kaufman, Statistical Decision and Related Techniques in Oil and Gas Exploration, (Prentice-Hall, 1963, New York City).
- [2] C. J. Grayson, Jr., Decisions Under Uncertainty: Drilling Decisions by and Gas Operators, (Division of Research, Harvard Business School, 1961, Boston).
- [3] J. J. Arps and T. G. Roberts, "Economics of Drilling for Cretaceous Oil Production on the East Flank of Denver- Julesburg Basin," Bull, AAPG, Vol. 42, No. 11, November, 1958.
- [4] Aitcheson, J. and Brown, J.A.C., The Lognormal Distribution, (Cambridge University Press, 1957, New York).

APPENDIX A

Mathematical Derivation of Plotting Method Used In Exhibits 1 through 9

Let \tilde{x} be a Lognormal random variable whose distribution function has parameter (μ, σ^2) so that

$$\tilde{x} \sim f_L(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2} \frac{1}{x}, \quad (1)$$

$$x > 0, \mu > 0, \sigma^2 > 0,$$

and

$$P(\tilde{x} < x) = \int_0^x f_L(z|\mu, \sigma^2) dz \equiv F_L(x|\mu, \sigma^2). \quad (2)$$

Then defining

$$u = [\log x - \mu]/\sigma, \quad (3)$$

we have

$$F_L(x|\mu, \sigma^2) = F_{N^*}(u) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{1}{2}v^2} dv. \quad (4)$$

Define u_i to be that value of u such that

$$F_{N^*}(u_i) = f_i \equiv \frac{N-n+i}{N+1}. \quad (5)$$

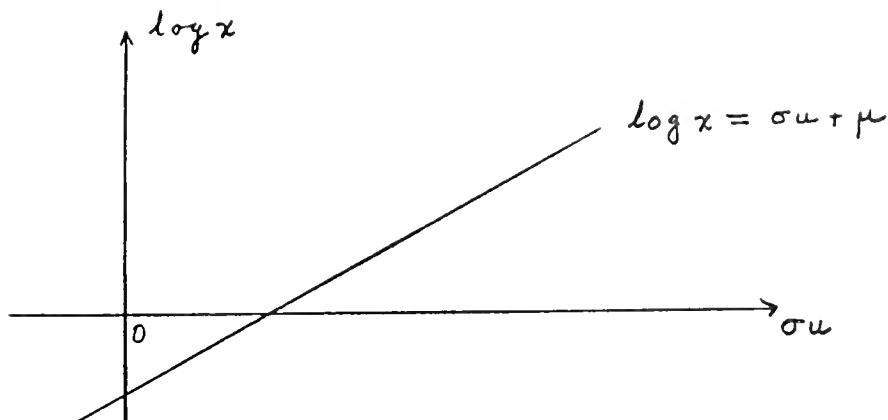
Let x_1, \dots, x_n be a sequence of n independent sample observations, each generated according to (1), ordered as to size ($x_i \geq x_j, n \geq i \geq j \geq 1$), and constituting the n largest sample observations from a sample of size $N > n$.

If we regard x_i as an estimate of the $(N-n+i/N+1)$ st fractile of the distribution of \tilde{x} , it follows from (3) and (5) that

$$\log x_i = \sigma u_i + \mu. \quad (6)$$

Given f_i we may compute u_i , from any table of the standardized Normal Cumulative Distribution Function.

Hence knowing x_i and f_i determines a linear relation between $\log x_i$ and u_i expressed in units of σ up to the additive constant μ ; e.g.



Given an ordered sequence x_1, \dots, x_n of sample observations we may display x_i , f_i , and u_i for $i=1, 2, \dots, n$ as below:

<u>Units u_i of σ</u>		
$\frac{x_i}{f_i}$	$\frac{f_i}{f_i}$	$\frac{u_i}{u_i}$
x_1	$N-n+1/N+1$	u_1
x_2	$N-n+2/N+1$	u_2
\vdots	\vdots	\vdots
x_n	$N/N+1$	u_n

We may plot ordered pairs $(u_i, \log x_i)$ on a graph such as that shown above, and fit a straight line to the plotted points. If in fact the x_i are generated according to (1) and n is "large" then the straight line should fit the data well, for the $(u_i, \log x_i)$ must satisfy (6). The intercept of the line with the vertical axis yields a graphical estimate of μ . The slope of the line yields a graphical estimate of σ . The graphical estimates displayed in Exhibit 9 were determined in this fashion.

APPENDIX B

Derivation of Partial Moments of the Probability Distribution of Reserves Remaining to be Discovered

From Assumption I and II we have

$$\tilde{r}_t \sim f_L(r_t | \mu_t, \sigma_t^{-2})$$

and

$$\tilde{s} \sim f_L(s | \mu_s, \sigma_s^{-2}) .$$

Using I and II, we derived the probability distribution (1) of \tilde{z} , reserves remaining to be discovered:

$$\tilde{z} \sim \kappa [f_L(z | \mu_s, \sigma_s^{-2}) - \frac{R_t}{S} f_L(z | \mu_t, \sigma_t^{-2})] . \quad (1)$$

The k th partial moment about the origin of \tilde{z} is easily found using formula (7.8b) of [1]:

$$E_0^a(\tilde{z}^k) \equiv \int_0^a y^k f_z(y) dy = \kappa e^{k \mu_s + \frac{1}{2} k^2 \sigma_s^2} F_{N^*}(w_1(a)) - \kappa \frac{R_t}{S} e^{k \mu_t + \frac{1}{2} k^2 \sigma_t^2} F_{N^*}(w_2(a)) \quad (2)$$

where

$$w_1(a) = [\log a - \mu_s] \sigma_s^{-1} - k \sigma_s$$

and

$$w_2(a) = [\log a - \mu_t] \sigma_t^{-1} - k \sigma_t , \quad \text{and}$$

where for notational simplicity we let $f_z(\cdot)$ represent the probability density function (1).

Proof: Observe that

$$\begin{aligned} E_0^a(\tilde{z}^k) &= \kappa \int_0^a y^k [f_L(y | \mu_s, \sigma_s^{-2}) - \frac{R_t}{S} f_L(y | \mu_t, \sigma_t^{-2})] dy \\ &= \kappa \int_0^a y^k f_L(y | \mu_s, \sigma_s^{-2}) dy - \kappa \frac{R_t}{S} \int_0^a y^k f_L(y | \mu_t, \sigma_t^{-2}) dy. \end{aligned}$$

Then (2) follows directly from (7.8b) of [1].

It follows that the formula for the incomplete kth moment of \tilde{z} from a to ∞ is

$$E_a^\infty(\tilde{z}^k) = \kappa e^{k\mu_s + \frac{1}{2}k^2\sigma_s^2} G_{N^*}(w_1(a)) - \kappa \frac{R_t}{S} e^{k\mu_t + \frac{1}{2}k^2\sigma_t^2} G_{N^*}(w_2(a))$$

where

$$G_{N^*}(\cdot) = 1 - F_{N^*}(\cdot) \quad .$$



BASEMENT

Date Due

[REDACTED]

JAN 03 '77

OCT 27 '77

NOV 29 '78

JUN 26 '79

SEP 11 '79

~~JAN 27 '80~~

JUL 7 1986

FEB 14 1990

MAY 1990

Lib-26-67



55-64



56-64



57-64

1751840

HD28

M.I.T. Alfred P. Sloan
School of Management
Working Papers.

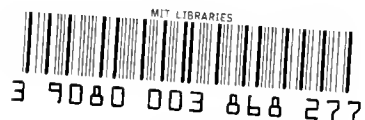
.M414

Nos. 55-64

1961



62-64



65-64



66-64



67-64



70-64



75-64



76-64



77-64



81-64



84

